

ODP-81-1224
18 SEP 1981

MEMORANDUM FOR: Chief, Public Affairs Branch

FROM:

[REDACTED]
Acting Director of Data Processing

SUBJECT:

Response to Public Affairs' Request for
DDA/ODP Assistance

REFERENCES:

- A. Memo to DDA from D/PA (DDA-81-1226),
dtd. 9 June 1981, SUBJECT: PRB
Reference Center
- B. Memo to D/PA from DDA (ODP-81-7058),
Same Subject

1. As agreed to in our 9 June 1981 meeting and documented in the referenced memoranda, a preliminary study of the Publication Review Board's information storage and retrieval needs has been completed. The attached paper contains the findings and recommendations of [REDACTED] and [REDACTED]

2. Their recommendation of a formatted file approach over a full text retrieval system would significantly reduce the resources required for converting textual manuscripts to machine readable form. Furthermore, effective indexing and abstracting will provide the retrieval flexibility needed by PRB. There is also a continuing resource implication to PAB for a formatted file. An operational system will require one full-time professional, as a data base manager/indexer. This professional would have to be provided from your staff. It will be difficult to recruit any individual with these skills below the GS-12/11 level. In addition, more indexing resources would be needed if you plan to convert the existing data base. D/OCR informs me that he does not have indexing personnel available for loan to PAB for this project.

3. The next step, the file design/requirements study, will require about three work months -- for an indexer, a computer systems specialist, and someone from your staff. Consequently, I hesitate to recommend such a step unless you feel confident that you have the necessary resources available for an operational system. I will await your response. Meanwhile, if you have any questions regarding the Preliminary Investigation Report, please call Mr. [REDACTED]

Preliminary Investigation Report
for the Publication Review Board

Prepared by

STAT
ODP Applications

and

STAT
OCR/ISG

September 10, 1981

1. Problem Definition - This preliminary investigation was conducted to determine what approach should be taken in providing an automated system for the storage and retrieval of pertinent information related to Publication Review Board's (PRB) pre-publication review process. The problem as stated by Office of Public Affairs (now Public Affairs Branch) is one of being able to recall what information has been disclosed to the general public through the review mechanism and what information has been withheld.

2. Findings - To begin, we believe that the PRB application is a good candidate for ADP control. The variety and amount of information to be controlled and the need for a timely, systematic organized search and retrieval apparatus supports this belief. Our initial reaction is that it is not a likely candidate for full text processing. Data conversion requirements, the size of the data base to be initially converted (40,000 pages), the projected file growth and storage requirements are the primary reasons for our decision. Eliminating full text processing as an alternative narrows the selection to a formatted file approach, that is, the creation of indexes/records containing information about the manuscripts; the manuscripts themselves being retained in a separate collection.

From a systems point-of-view, the consideration of a formatted file application brings up many points regarding support of the application that should be addressed before a decision to proceed is made. Such an approach will require considerable resources for data reduction, input and file maintenance. It will require a disciplined environment that includes an information abstraction and data entry capability as well as a quality control mechanism. Additionally it could introduce complexities and changes in PRB's office procedures and responsibilities that could affect system design. For example, procedures may have to be established for logging and tracking the manuscript in order to insure that the final disposition has been made and the file record is complete.

In order to assist PRB in analyzing their needs and commitments we have constructed a file resources strawman (attachments 1-5). These estimates are based on a review of a sample of manuscript files currently held in PRB and from initial discussions with PRB personnel.

Using the attached estimates we recommend at least one person fulltime to support current file needs. This estimate presumes this person will have the various skills necessary to perform the functions of control, abstract, input, maintain and retrieve, and a first hand awareness of on-line data entry and ad hoc subject retrieval. Ideally a fully trained and experienced abstractor/indexer would be

desirable. This experience is absolutely necessary to initially maintain the lower range of time estimates and to support a high retrieval/indexing relevancy rate. Of particular concern is the time allocated to data base management functions. At implementation this expenditure will be weighted to the high range figure. Gradually as experience grows and as reference tools are completed the expenditure should ease. About six months will be required for this cycle to settle down.

In addition to keeping up with current receipts the conversion of present file holdings is recommended. The conversion of this data base is estimated to require approximately 1/2 manyear. Using the lower resource allocation figure we anticipate this task to complement and to support the current file building operations. It of course will slow down this process unless additional resources are allocated.

The strawman record structure is based on three groups of information about a manuscript - bibliographic data, an abstract of the theme and/or subjects treated and an abstract of the reviewers' comments. Each information group has been described as a subrecord. These subrecords are considered, for the purpose of this file estimate, to be independent for input and maintenance activities. That is, each subrecord may be input to the system as it is completed rather than delaying input until all subrecords are available. Intermittent input allows the system to serve as a control and tracking tool as well as a retrospective retrieval device. Special emphasis on maintenance functions is stressed as each subrecord may be accessed several times to input information as it becomes available; this is especially true in subrecords 1 and 3. At retrieval, however, the record is addressed as a coordinated whole.

3. Recommendations - If based on these data a decision to proceed is made, we would then recommend the formation of a file design team. Composed of a PRB representative, a computer system analyst, and an indexing expert, this team would be responsible for a complete system requirements and file design document. After the requirements have been defined, the group will dissolve and the ODP analyst will write a project proposal for a system to be developed by ODP Applications. This proposal will include all aspects of system design, development, and implementation.

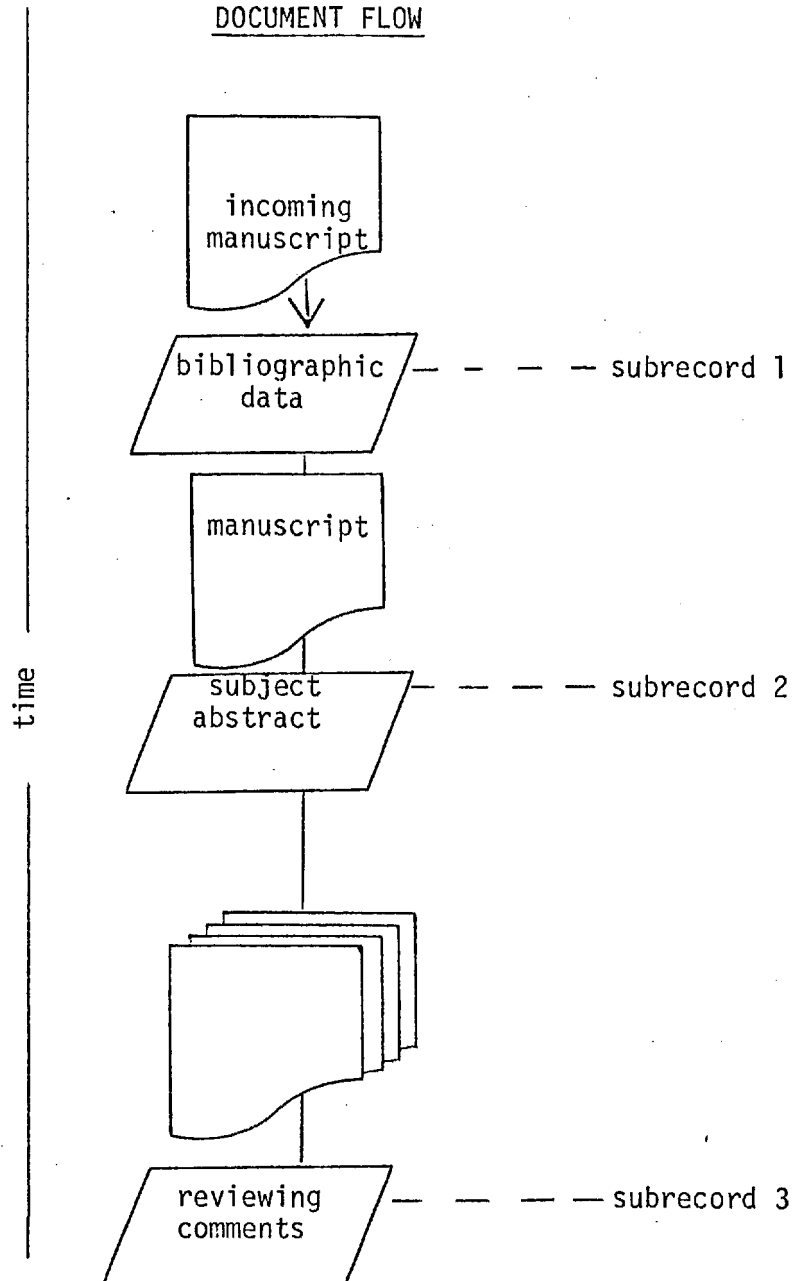
PRB FILE "STRAWMAN"

RECORD STRUCTURE

Each manuscript is represented by one-three part index record. A record is not complete until all three subrecords are input. Each subrecord, however, may be input separately in a unique maintenance action.

Subrecord 1 contains	bibliographic data examples: author's name, title, PRB control number, date submitted, document type, date of comments	comments: basically data data currently controlled in a PRB RAMIS formatted file -- with certain standardizations. (dates, document type, name)
	estimated size 400 characters	
Subrecord 2 contains	subject abstract (keywords/ keyword phrases) examples: media control, disinforma- tion, CIA field station	comment: this strawman uses keywords/keyword phrases with- out additional encoding. The use of codes to represent concepts and/or areas should be considered in future requirements studies. In addition the linkage of areas to keywords/concepts is viewed as a necessary retrieval requirement.
STAT	<div style="border: 1px solid black; height: 80px; width: 100%;"></div>	
	estimated size 750 characters	
Subrecord 3 contains	reviewing official's com- ments and/or concerns (key- words/keyword phrases) examples: operations -	comment: this strawman uses keywords/keyword phrases with- out additional encoding. The use of codes to represent concepts and/or areas should be considered in future require- ments studies. As in subrecord 2 the linkage of area with the keywords/concepts is most im- portant. The addition of page number to the indexing phrase is an enhancement that may have merit.
	estimated size 750 characters	

PRB FILE "STRAWMAN"

DOCUMENT FLOW

DATA BASE SIZE AND GROWTH RATE

PRB holdings as of 1 August 1981

109 books
270 articles
21 book reviews
11 outlines
12 speeches
27 other
450

Distributed Record Size

	Greater Manuscripts (Books)	Lesser Manuscripts (Articles, etc.)
Subrecord 1	400 char.	400 char.
Subrecord 2	750 char.	350 char.
Subrecord 3	<u>750</u> char. 1,900 char.	<u>350</u> char. 1,100 char.

Data Base Size - to be converted
(pre CY Aug 81)

Books 109 x 1,900 char. = 207,100 char.
Articles, etc. 341 x 1,100 char. = 375,100 char.
TOTAL = 582,200 char.

Growth Rate (based on projected
CY 81 rate)

Books 24 x 1,900 char. = 45,600 char.
Articles, etc. 176 x 1,100 char. = 193,600 char.
TOTAL = 239,200 char.

PRB FILE "STRAWMAN"

PRB RESOURCES REQUIRED FOR CURRENT DATA BASE MANAGEMENT
 (based on projected CY81 input rate)

<u>FUNCTION</u>	<u>TYPE OF MANUSCRIPT</u>	<u>TIME REQUIRED/ MANUSCRIPT</u>	<u>X</u>	<u>RATE OF INPUT/YEAR</u>	<u>TOTAL TIME/YEAR</u>
Bibliographic Indexing	Books	15-30 min		24	6 - 12
	Articles	15-30 min		176	44 - 88
Abstracting - Subject	Books	2-4 hrs		24	48 - 96
	Articles	30 min - 1 hr		176	88 - 176
Abstracting - Index reviewers' Comments	Books	2-4 hrs		24	48 - 72
	Articles	15-30 min		176	44 - 88
Data Entry	Books	30 min - 1 hr		24	12 - 24
	Articles	15-30 min		176	44 - 88
Data Base Mgt-	---	2-3 hrs/day		---	520 - 780
TOTAL					854 - 1,424 manhours

PRB FILE "STRAWMAN"

RESOURCES FOR DATA BASE CONVERSION
 (based on current holdings)

<u>FUNCTION</u>	<u>TYPE OF MANUSCRIPT</u>	<u>TIME REQUIRED/ MANUSCRIPT</u>	X	<u>NUMBER CURRENTLY HELD BY PRB</u>	<u>TOTAL HOURS</u>
Bibliographic Indexing	Books	15 min		109	27.25
	Articles	15 min		341	85.25
Abstracting - Subject	Books	2 hrs		109	218
	Articles	30 min		341	170.50
Abstracting - Index Reviewers' Comments	Books	2 hrs		109	218
	Articles	15 min		341	85.25
Data Entry	Books	30 min		109	54.50
	Articles	15 min		341	85.25
TOTAL					944. manhours